The Cost of Interference in Evolving Multiagent Systems (Extended Abstract)

The Anh Han
Tesside University
t.han@tees.ac.uk

Long Tran-Thanh
University of Southampton
l.tran-thanh@soton.ac.uk

Nicholas R. Jennings
University of Southampton
nrj@ecs.soton.ac.uk

ABSTRACT

We study the situation of a decision-maker who aims to encourage the players of an evolutionary game theoretic system to follow certain desired behaviours. To do so, she can interfere in the system to reward her preferred behavioural patterns. However, this action requires certain cost (e.g., resource consumption). Given this, her main goal is to maintain an efficient trade-off between achieving the desired system status and minimising the total cost spent. Our results reveal interesting observations, which suggest that further investigations in the future are required.

1. INTRODUCTION

In this paper we consider the following problem. Given a system with a finite number of players, who interact with each other either repeatedly or in a one-shot manner. A decision-maker, who is not part of the system, aims to force the players to maintain certain strategy profiles. However, the decision-maker does not fully control all the behaviours and actions of the players, due to some (physical) limitations. Instead, she can interfere in the system at any particular time step, (partially) modifying the system dynamics. By doing so, she has to consume a certain amount of her (typically limited) resources, which is an increasing function of the degree of interference. Given this, the research challenge is to identify a sequence of actions that balances between achieving the decision-maker's objective (i.e., to maintain a desired state) and minimising the resource consumption. This model is motivated by many real-world applications, such as the peace-keeping process of the United Nations, or population control in habitat management. Although the (sequential) decision-making literature provides a number of techniques to tackle similar resource constraint optimisation problems [4], these approaches typically ignore the fact that the players, with whom the decision-maker has to interact, also have their own strategic behaviours that together drive the dynamics of the system. Given this, we argue that such solutions will not be able to exploit the system characteristics, and thus, will fail in providing efficient performance in achieving the desired goals. On the other hand, game theoretic literature typically focusses on the extremes. In particular, researchers either assume that the system is fully closed (i.e., there is no outsider decision-makers), or the decision-maker has a full control on the behaviour of the players. Typical models for the former are classical (both non-cooperative and coalitional) game theoretical models. The latter includes models from mechanism design, where the decision-maker is the system designer, and can define some set of norms and penalties such that the players are not incentivised to deviate from the norms.

Against this background, this paper aims at filling the gap by addressing the problem as follows. We combine the decision-making process design with an evolutionary game theoretic perspective (described in Section 2). While the former aims at capturing the behaviour of the decision-maker, the latter can be used to formalise the dynamics of the system of players. In particular, we consider a population where the players interact through the Prisoner's Dilemma. Suppose that as an outsider decision-maker, we aim to promote a certain strategy profile. We also have a budget that can be used to interfere by rewarding particular strategists/individuals in the population at concrete moments (e.g. depending on the current composition of the population). In particular, at each time step, we can reward the players who follow the desired strategy. Hence, the research question here is to identify when and how much we want to pay the players, in order to achieve our goals.

2. MODEL AND METHODS

We focus here on a two-player game model, where the one-shot Prisoner's Dilemma (PD) [3] is used as the interaction model of agents in a population. The PD game is a well-known framework to study the problem of the evolution of cooperation [3], where without any supporting mechanisms such as kin selection, reciprocities, structured population, punishment and reward [2] and commitments [1], cooperation is rare and cannot evolve. Here, differently from previous work, we study what are the appropriate interference strategies (by rewarding cooperation) leading to high levels of cooperation while minimising the investment budget.

In a PD, a player can choose either to cooperate (C) or defect (D). A player who chooses to cooperate with someone who defects receives the sucker's payoff S, whereas the defecting player gains the temptation to defect, T. Mutual cooperation (resp., defection) yields the reward R (resp., punishment P) for both players. The PD is characterized by the ordering, $T > R > P > S$, where in each interaction defection is the rational choice but cooperation is the desired outcome.

In addition, we consider a well-mixed population of N individuals. The individuals adopt one of the two pure strategies: C (always cooperates) or D (always defects). In a population with k C-players and $(N-k)$ D-players, the average payoff a C- and a D-player can be written as follows, respectively: $\Pi_C(k) = \frac{1}{N-1}\sum_{j=1}^{m}[(k-1)R + (N-k)S]$; $\Pi_D(k) = \frac{1}{N-1}\sum_{j=1}^{m}[(N-k-1)T + kP]$.

In our model, we adopt a standard approach to implementing social learning or imitation [3]. Namely, at each time-step, one individual i with a fitness f_i is randomly chosen for behavioural

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.*

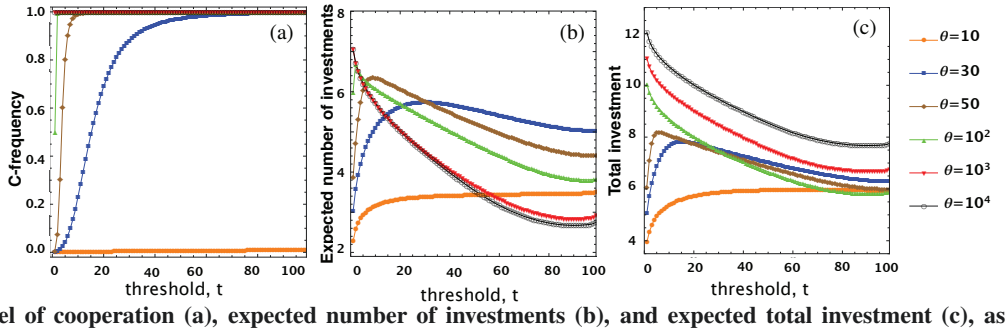

Figure 1: Level of cooperation (a), expected number of investments (b), and expected total investment (c), as functions of the investment threshold t and for different per-generation investment values θ. In panel (b) and (c), the plot is on a log(10)-scale. Parameters: $R = 1,\ T = 2, P = 0,\ S = -1$; population size $N = 100$; imitation strength $\beta = 0.1$.

revision. i will adopt the strategy of a randomly chosen individual j with fitness f_j with a probability given by the Fermi function $\left(1 + e^{-\beta(f_j - f_i)}\right)^{-1}$, where the quantity β controls the intensity of selection. Furthermore, we adopt the small mutation approach, i.e. a single mutant in a monomorphic population will fixate or will become extinct long before the occurrence of another mutation. This allows one to describe the evolutionary dynamics of our population in terms of a reduced Markov Chain of a size equal to the number of different strategies. The stationary distribution of the Markov Chain characterises the average time the population spends in each of these monomorphic states and can be computed analytically. Due to lack of space, we refer to the Method session in [1] for a full description of the stationary distribution computation.

We now define the optimization problem. We consider that the investment strategy solely depends on the current state of the population. Namely, whenever there are i C-players (i.e. $N - i$ D-players) in the population, an (per-generation) investment, θ_i, is made. That is, each C-player gets an increase of θ_i/i in the average payoff. In order to compute the expected total amount of investment we need to compute the expected number of times the population contains i C-players, $1 \leq i \leq N-1$. For that, we consider an absorbing Markov chain of $(N+1)$ states, $\{S_0, ..., S_N\}$, where S_i represents a population with i C-players. S_0 and S_N are absorbing states. Let $U = \{u_{ij}\}_{i,j=1}^{N-1}$ denote the transition matrix between the $N-1$ transient states, $\{S_1, ..., S_{N-1}\}$. The transition probabilities can be defined as follows. For $1 \leq i \leq N-1$,

$$u_{i,i\pm j} = 0 \qquad \text{for all } j \geq 2$$
$$u_{i,i\pm 1} = \frac{N-i}{N}\frac{i}{N}\left(1 + e^{\mp\beta[\Pi_C(i) - \Pi_D(i) + \theta_i/i]}\right)^{-1}$$
$$u_{i,i} = 1 - u_{i,i+1} - u_{i,i-1}$$

The entries n_{ij} of the so-called fundamental matrix $N = (I - U)^{-1}$ of the absorbing Markov chain gives the expected number of times the population is in the state S_j if it is stated in the transient state S_i. As a mutant can randomly occur either at S_0 or S_N, the expected number of visits at state S_i is: $\frac{1}{2}(n_{1i} + n_{N-1,i})$. Hence, the expected total investment is: $Q = \frac{1}{2}\sum_{i=1}^{N-1}(n_{1i} + n_{N-1,i})\theta_i$.

In short, the goal is to find an investment strategy that maximizes the cooperation level (or guarantees a certain level of cooperation) while minimising the expected total investment Q.

3. NUMERICAL EVALUATION

We analyze a concrete investment strategy, supported by real world scenarios, in which we have a fixed amount of resource, θ, for rewarding cooperative acts in each generation, i.e. $\theta_i = \theta\ \forall i$. We ask, should one focus the effort to reward only a few C players rather than spreading the effort to reward all C players but that may not be sufficient for them to survive? For that, we consider investment strategies that invest only when the number of C-players does not exceed a given threshold t, $1 \leq t \leq N-1$.

In Figure 1a, we plot the frequency (level) of cooperation varying the investment threshold t, and for different per-generation investment θ. It is not surprising that the larger threshold t, i.e. the more spreading the investment, and the larger the per-generation investment (θ), the higher level of cooperation is obtained. For a too small θ, defection is prevalent even when the investment is always made (see $\theta = 10$). For a sufficiently large θ, a rather spreading investment strategy can lead to a high level of cooperation. But does a more spreading investment scheme necessarily mean a larger amount of total investment, let alone the higher level of cooperation it leads to? If the stochastic and dynamic aspects of the system are not taken into account, the answer is clearly the positive one. However, as one can see from Figure 1c where the expected total investment is shown for varying t, above a certain threshold of t, a more spreading investment strategy mostly leads to a lower total investment expected to be made (for $\theta \geq 30$). Moreover, the larger θ is, the lower that threshold and the more significant the decreasing are. But it is important to note that this decreasing tendency stops when t reaches a certain threshold (then it slowly increases) (e.g. for $\theta = 100$, the optimal is $t = 91$, and for $\theta = 50$, the optimal is $t = 97$). The explanation for this observation can be seen from Figure 1b, where the expected number of times of investment is depicted for varying t. We can see that it is still important to make investments when there is a rather large fraction of C-players (i.e. high enough value of t) in the population, because otherwise defection can still fight back and becomes more frequent, leading to further investments latter (hence, wasting the earlier investment efforts). The investment can be ceased only when the fraction of C-players in the population is sufficiently large (around 90%) to be able to maintain their abundance themselves. That is, once we decide to interfere (to help with sustaining high cooperation), we should interfere until the cooperators can survive and fight defection on their own.

Additional analysis shows that our results are robust for varying different parameters, including θ and the payoff matrix entries.

4. REFERENCES

[1] T.A. Han, L.M. Pereira, F.C. Santos, and T. Lenaerts. Good agreements make good friends. *Scientific reports*, 3(2695), 2013.

[2] K Sigmund, C Hauert, and M Nowak. Reward and punishment. *PNAS*, 98(19):10757–10762, 2001.

[3] Karl Sigmund. *The Calculus of Selfishness*. Princeton University Press, 2010.

[4] L. Tran-Thanh, A. Chapman, A. Rogers, and N. R. Jennings. Knapsack based optimal policies for budget–limited multi–armed bandits. *AAAI*, pages 1134–1140, 2012.